Call for papers: PDSW'23

The 8th International Parallel Data Systems Workshop
http://www.pdsw.org/

November 12, 2023 1:30 PM - 5:00 PM

Held in conjunction with SC23, Denver, CO

General Chair
Amelie Chi Zhou, Shenzhen University, China

Program Co-Chairs
Bing Xie, Microsoft, USA
Suren Byna, The Ohio State University, USA

Reproducibility Co-Chairs
Tanu Malik, DePaul University, USA
Jean Luca Bez, Lawrence Berkeley National Laboratory, USA

Publicity Chair
Kira Duwe, EPFL, Switzerland

Web and Proceedings Chair
Joan Digney, Carnegie Mellon University

Important Dates

Regular Papers and Reproducibility Study Papers
Submissions due: Aug 6th, 2023, 11:59 PM AoE
Paper Notification: Sept 8th, 2023, 11:59 PM AoE
Camera ready due: Sept 29th, 2023, 11:59 PM AoE

Work in Progress (WIP)
Submissions due: Sept 15th, 2023, 11:59PM AoE
WIP Notification: On or before Sept 23nd, 2023

Workshop Abstract:

We are pleased to announce the 8th International Parallel Data Systems Workshop (PDSW'23). PDSW'23 will be hosted in conjunction with SC23: The International Conference for High Performance Computing, Networking, Storage and Analysis, in Denver, CO.

Efficient data storage and data management are crucial to scientific productivity in both traditional simulation-oriented HPC environments and Big Data analysis environments. This issue is further exacerbated by the growing volume of experimental and observational data, the widening gap between the performance of computational hardware and storage hardware, and the emergence of new data-driven algorithms in machine learning. The goal of this workshop is to facilitate research that addresses the most critical challenges in scientific data storage and data processing. PDSW will continue to build on the successful tradition established by its predecessor workshops: the Petascale Data Storage Workshop (PDSW, 2006-2015) and the Data Intensive Scalable Computing Systems (DISCS 2012-2015) workshop. These workshops were successfully combined in 2016, and the resulting joint workshop has attracted up to 38 full paper submissions and 140 attendees per year from 2016 to 2022.

We encourage the community to submit original manuscripts that:

- introduce and evaluate novel algorithms or architectures,
- inform the community of important scientific case studies or workloads, or
- validate the reproducibility of previously published work

Special attention will be given to issues in which community collaboration is crucial for problem identification, workload capture, solution interoperability, standardization, and shared tools. We also strongly encourage papers to share complete experimental environment information (software version numbers, benchmark configurations, etc.) to facilitate collaboration.

Topics of interest include the following:

- Large-scale data caching architectures
- Scalable architectures for distributed data storage, archival, and virtualization
- The application of new data processing models and algorithms towards computing and analysis
- Performance benchmarking, resource management, and workload studies
- Enabling cloud and container-based models for scientific data analysis
- Techniques for data integrity, availability, reliability, and fault tolerance
- Programming models and big data frameworks for data intensive computing
- Hybrid cloud/on-premise data processing
- Cloud-specific data storage and transit costs and opportunities
- Programmability of storage systems
- Data filtering, compression, reduction techniques
- Data and metadata indexing and querying
- Parallel file systems, metadata management, and complex data management
- Integrating computation into the memory and storage hierarchy to facilitate in-situ and in-transit data processing
- Alternative data storage models, including object stores and key-value stores
- Productivity tools for data intensive computing, data mining, and knowledge discovery
- Tools and techniques for managing data movement among compute and data intensive components
- Cross-cloud data management
- Storage system optimization and data analytics with machine learning
- Innovative techniques and performance evaluation for new memory and storage systems

Regular Paper Submissions

All papers will be evaluated by a competitive peer review process under the supervision of the workshop program committee. Selected papers and associated talk slides will be made available on the workshop web site. The papers will also be published in the SC23 Workshop Proceedings.

Authors of regular papers are strongly encouraged to submit Artifact Description (AD) Appendices that can help to reproduce and validate their experimental results. While the inclusion of the AD Appendices is optional for PDSW'23, submissions that are accompanied by AD Appendices will be given favorable consideration for the *PDSW Best Paper award*.

PDSW'23 follows the [SC23 Reproducibility Initiative](). For Artifact Description (AD) Appendices, we will use the format of the SC23 for PDSW'23 submissions. The AD should include a field for one or more links to data (zenodo, figshare, etc.) and code (github, gitlab, bitbucket, etc.) repositories. For the Artifacts that will be placed in the code repository, we encourage authors to follow the [PDSW 2023 Reproducibility Addendum]() on how to structure the artifact, as it will make it easier for the reviewing committee and readers of the paper in the future.

For PDSW 2023, we WILL NOT be taking applications for badges or awarding them due to time constraints. We will still provide reviews and feedback on the ADs, but AE will not be reviewed and badges will not be awarded.

Submit a not previously published paper as a PDF file, indicate authors and affiliations. Papers must be up to 6 pages, not less than 10 point font and not including references and optional reproducibility appendices.

Submission site: https://submissions.supercomputing.org/

Submissions due: Aug 6th, 2023, 11:59 PM AoE

Papers must use the ACM conference paper template available at:
https://www.acm.org/publications/proceedings-template

Work in progress (WIP) Session

There will be a WIP session where presenters provide brief 5-minute talks on their on-going work, with fresh problems/solutions. WIP content is typically material that may not be mature or complete enough for a full paper submission and will not be included in the proceedings. A one-page abstract is required. Submission site: https://submissions.supercomputing.org/